Application of Information Theory, Lecture 5

Channel Capacity and Isoperimetric Inequality

Handout Mode

Iftach Haitner

Tel Aviv University.

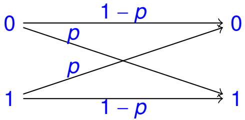
November 25, 2014

Part I

Channel Capacity

The problem

- We want to send a message $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$, but the communication channel is faulty
- Each bit is (independently) flipped w.p. p (e.g., 0.1)



- (expected) Error rate is p
- Such "channel" is called Binary Symmetric Channel (BSC)
- When sending m bits, we have $\approx pm$ errors
- Can we send bits with smaller error?

Solution

- Obvious solution is “error correction codes (ECC)”
- Most simple example: send each bit three times, and take majority
- Error happens if the channel errs at least twice
- For $p = 0.1$: happens w.p. $3p^2(1-p) + p^3 = 3 \cdot 0.01 \cdot 0.9 + 0.001 = 0.028$
- Error rate: $.028$
- Transmission rate: $1/3$ (i.e., # of bits recovered / #of bits transmitted)
- We reduced the error rate by reducing the transmission rate.
- Can we reduce the error rate, without reducing the transmitting rate too much?
- Before Shannon it was believed that very small error rate requires very small transamination rate.

Shannon's result

- Shannon showed that you can reduce the error rate towards 0, without reducing the transmission rate towards 0
- For any $c < C_p$, exists a code with transmission rate c that is correct w.p.
- Example: for $p = .1$, $C_p > \frac{1}{2}$.

  Hence, for sending $\mathbf{x} = (x_1, \ldots, x_m)$, one can send $2m$ bits, such that $\mathbf{x}$ is recovered w.p. close to 1
- More generally, $\forall p \in [0,1] \quad \exists C_p$ such that for sending $\mathbf{x} \in \{0,1\}^m$, one can send $\approx \frac{m}{C_p}$ bits, and $\mathbf{x}$ is recovered w.p. close to 1
- C_p might be 0 (i.e., for $p = \frac{1}{2}$)
- A revolution in EE and the whole world

Error correction code

- Message to send $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$
- Encoding scheme: $f\colon \{0,1\}^m \mapsto \{0,1\}^n$ $\qquad (n > m)$
- Decoding scheme: $g\colon \{0,1\}^n \mapsto \{0,1\}^m$
- $\frac{m}{n}$ — transmission rate
- Sender sends $f(x)$ rather than x
- Receiver decodes the message by applying g
- $\underbrace{\mathbf{x}}_{m \text{ bits}} \xrightarrow{\text{encoding}} \underbrace{f(\mathbf{x})}_{n \text{ bits}} \xrightarrow{\text{channel}} \underbrace{f(\mathbf{x}) \oplus Z}_{\text{bitwise XOR}} \xrightarrow{\text{decoding}} g(f(\mathbf{x}) \oplus Z)$

  $Z = (Z_1, \ldots, Z_n)$ where $Z_1, \ldots, Z_n$ iid $\sim (1-p, p)$ (i.e., over $\{0,1\}$ with $\Pr[Z_i = 1] = p$)
- We hope $g(f(\mathbf{x}) \oplus Z) = \mathbf{x}$
- ECCs are everywhere
- ECC Vs compression

Shannon's theorem

Theorem 1

$\forall p \quad \exists C_p, \ s.t. \ \forall \varepsilon > 0 \quad \exists m_\varepsilon, \ s.t. \ \forall m > m_\varepsilon$ *and* $n > m(\frac{1}{C_p} + \varepsilon)$,
$\exists \ f \colon \{0,1\}^m \mapsto \{0,1\}^n$ *and* $g \colon \{0,1\}^n \mapsto \{0,1\}^m, \ s.t. \ \forall \boldsymbol{x} \in \{0,1\}^m$:

$$\Pr_{z \leftarrow Z = (Z_1,\ldots,Z_n)} [g(f(\boldsymbol{x}) \oplus z) \neq \boldsymbol{x}] \leq \varepsilon$$

for $Z_1, \ldots, Z_n$ *iid* $\sim (1-p, p)$.

- $C_p = 1 - h(p)$ — the channel capacity

  $p = .1 \implies C_p = 0.5310 > \frac{1}{2}$

  $p = .25 \implies C_p \approx \frac{1}{5}$
- Tight theorem
- We prove a weaker variant that holds w.h.p. over $\mathbf{x} \leftarrow \{0,1\}^m$

Hamming distance

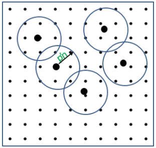
- For $\mathbf{y} = (y_1, \ldots, y_n) \in \{0,1\}^n$, let $|\mathbf{y}| = \sum_i y_i$ — Hamming weight of $\mathbf{y}$
- $|\mathbf{y} - \mathbf{y}'| = |\mathbf{y} \oplus \mathbf{y}'|$ — Hamming distance of $\mathbf{y}$ from $\mathbf{y}'$; # of places differ.

Proving the theorem

- Fix $p \in [0, \frac{1}{2})$ and $\varepsilon > 0$, and let $m > m_\varepsilon$ and $n \geq m(\frac{1}{C_p} + \varepsilon)$, for m_ε to be determined by the analysis. (Recall $C_p = 1 - h(p)$).
- We show $\exists\ f\colon \{0,1\}^m \mapsto \{0,1\}^n$ and $g\colon \{0,1\}^n \mapsto \{0,1\}^m$, s.t. $\Pr_{\mathbf{x} \leftarrow \{0,1\}^m} [g(f(\mathbf{x}) \oplus Z) \neq \mathbf{x}] \leq \varepsilon$
- $g(y)$ returns $\mathrm{argmin}_{\mathbf{x}' \in \{0,1\}^m} |y - f(\mathbf{x}')|$
- So it all boils down to finding f s.t. $\Pr_{\mathbf{x} \leftarrow \{0,1\}^m ; y = f(\mathbf{x}) \oplus Z} [\forall \mathbf{x}' \in \{0,1\}^m \setminus \{\mathbf{x}\} \colon |f(\mathbf{x}) - y| < |f(\mathbf{x}') - y|] \geq 1 - \varepsilon$
- Idea: for $p' > p$ to be determined later, find f s.t. w.h.p. over $\mathbf{x}$ and Z:

  (1) $|f(\mathbf{x}) \oplus Z, f(\mathbf{x})| \leq p'n$

  (2) $|f(\mathbf{x}) \oplus Z, f(\mathbf{x}')| > p'n$ for all $\mathbf{x}' \neq \mathbf{x}$



- We choose f uniformly at random (what does it mean?)
- Non-constructive proof
- Probabilistic method

Proving there exists good f

- Fix $p' > p$ such that $\frac{1}{C_{p'}} - \frac{1}{C_p} \leq \frac{\varepsilon}{2}$
- For $y \in \{0,1\}^n$, let $B_{p'}(y) = \{y \in \{0,1\}^n \colon |y' - y| \leq p'n\}$

(1) By weak low of large numbers, $\exists n' \in \mathbb{N}$ s.t. $\forall n \geq n'$ and $\forall \mathbf{x} \in \{0,1\}^m$:
$\alpha_n := \Pr_{z \leftarrow Z}\left[(f(\mathbf{x}) \oplus z) \notin B_{p'}(f(\mathbf{x}))\right] \leq \frac{\varepsilon}{2}$ (for any fixed f)

- Fact (proved later): $b(p') = |B_{p'}(y)| \leq 2^{n \cdot h(p')}$

$\Longrightarrow$ $\forall \mathbf{x} \neq \mathbf{x}' \in \{0,1\}^m$: $\Pr_{f,Z}\left[f(\mathbf{x}) \oplus Z \in B_{p'}(f(\mathbf{x}'))\right] = \frac{b(p')}{2^n} \leq \frac{2^{n \cdot h(p')}}{2^n} = 2^{-nC_{p'}}$

$\Longrightarrow$ $\forall \mathbf{x} \in \{0,1\}^m$: $\Pr_{f,Z}\left[\exists \mathbf{x}' \neq \mathbf{x} \in \{0,1\}^m \colon f(\mathbf{x}) \oplus Z \in B_{p'}(f(\mathbf{x}'))\right] \leq 2^{m - nC_{p'}}$

$\Longrightarrow$ $\exists f$ s.t.
$\beta_{m,n} := \Pr_{\mathbf{x} \leftarrow \{0,1\}^m}\left[\exists \mathbf{x}' \neq \mathbf{x} \in \{0,1\}^m \colon f(\mathbf{x}) \oplus Z \in B_{p'}(f(\mathbf{x}'))\right] \leq 2^{m - nC_{p'}}$

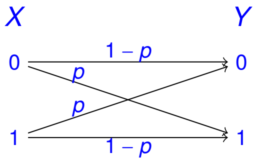
$\Longrightarrow$ $\beta_{m,n} \leq \frac{\varepsilon}{2}$, for $n \geq \frac{m}{C_{p'}} - \log \varepsilon + 1 = m\left(\frac{1}{C_{p'}} + \frac{1 - \log \varepsilon}{m}\right)$

(2) $\beta_{m,n} \leq \frac{\varepsilon}{2}$, for $m \geq m' = \frac{2(1 - \log \varepsilon)}{\varepsilon}$ and $n \geq m\left(\frac{1}{C_p} + \frac{\varepsilon}{2} + \frac{1 - \log \varepsilon}{m}\right) = m\left(\frac{1}{C_p} + \varepsilon\right)$

- Hence, for $m > m_\varepsilon = \max\{m', n'\}$ and $n > m\left(\frac{1}{C_p} + \varepsilon\right)$, it holds that
$\Pr_{\mathbf{x} \leftarrow \{0,1\}^m}\left[g(f(\mathbf{x}) \oplus Z) \neq \mathbf{x}\right] \leq \alpha_n + \beta_{m,n} \leq \varepsilon$. $\square$

Why $C_p = 1 - h(p)$?

- Let $X \leftarrow \{0,1\}$, $Z \sim (1-p,p)$ and $Y = X \oplus Z$



- $I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(Z) = 1 - h(p) = C_p$
- Received bit “gives" C_p information about transmitted bit
- Hence, to recover m bits, we need to send at least $m \cdot \frac{1}{C_p}$ bits
- A different proof:

Size of bounding ball

Claim 2

For $p \in [0, \frac{1}{2}]$ and $n \in \mathbb{N}$: it holds that $\sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k} \leq 2^{n \cdot h(p)}$

Proof in a few slides (we already saw that $\binom{n}{pn} \approx 2^{n \cdot h(p)}$)

Corollary 3

For $y \in \{0,1\}^n$ and $p \in [0, \frac{1}{2}]$, let $B_p(y) = \{y \in \{0,1\}^n : |y' - y| \leq pn\}$. Then $|B_p(y)| = \sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k} \leq 2^{n \cdot h(p)}$

Very useful estimation. Weaker variants follows by AEP or Stirling,

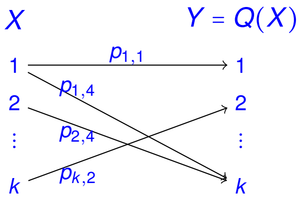
Tightness

- $X \leftarrow \{0,1\}^m$, $Z = (Z_1, \ldots, Z_n)$ where $Z_1, \ldots, Z_n$ iid $\sim (1-p, p)$
- $\underbrace{X}_{m \text{ bits}} \longrightarrow \underbrace{f(X)}_{n \text{ bits}} \longrightarrow \underbrace{f(X) \oplus Z}_{Y} \longrightarrow \underbrace{g(f(X) \oplus Z)}_{g(Y)}$
- Assuming $\Pr[g(Y) = X] \geq 1 - \varepsilon$, we show $nC_p \geq m(1-\varepsilon) - 1$
- Compare to $nC_p > m(1 + \varepsilon C_p)$ in Thm 1
- Hence, $\lim_{\varepsilon \to 0} \frac{m}{n} = C_p$
- By Fano, $H(X|Y) \leq h(\varepsilon) + \varepsilon \log(2^m - 1) \leq 1 + \varepsilon m$
- $I(X;Y) = H(X) - H(X|Y) \geq m - \varepsilon m - 1 = m(1-\varepsilon) - 1$
- $H(Y|X) = H(X,Y) - H(X) = H(X,Z) - H(X) = H(Z) = nh(p)$
- $I(X;Y) = H(Y) - H(Y|X) = n - nh(p)$
- Hence, $m(1-\varepsilon) \leq I(X;Y) + 1 = n(1 - h(p)) + 1 = nC_p + 1$
- Alternative proof

General communication channel

$Q \colon [k] \mapsto [k]$ that channel (a probabilistic function)

$p_{i,j} = \Pr[Q(i) = j]$



- $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$
- Encoding function $f \colon \{0,1\}^m \mapsto \{1, \ldots, k\}^n$
- Decoding function $g\{1, \ldots, k\}^n \mapsto \{0,1\}^m$
- $\mathbf{x} \xrightarrow{\text{encoding}} f(\mathbf{x}) \xrightarrow{\text{channel}} Q(f(\mathbf{x})) \xrightarrow{\text{decoding}} g(Q(f(\mathbf{x})))$
- We hope for $g(Q(f(\mathbf{x}))) = \mathbf{x}$
- Channel capacity $C_Q = \max_X I(X;Y)$
- The maximal information Y gives on X
- Shannon theorem: $\forall Q$ and $\forall \varepsilon > 0$, $\exists m_\varepsilon \colon \forall m > m_\varepsilon$ and $\forall n > m(\frac{1}{C_Q} + \varepsilon)$: $\exists f, g$ as above s.t. $\Pr_Q[g(Q(f(\mathbf{x}))) \neq \mathbf{x}] \leq \varepsilon$, for all $\mathbf{x} \in \{0,1\}^m$.
- Proof: similar lines to the binary case, but more subtle distribution for f

Discussion

- Tight result
- Non-constructive
- Coding theory: design explicit (and efficient) code achieving the same bounds
- Application: faulty communication, storage
- Combination of data compression and ECC

Part II

Combinatorial Applications

Movies

- 2^n people, $m = 3n$ movies.
- Every pair of movies was seen by at least 90% of the people
- Claim: there exist two people who saw exactly the same set of movies
- $X \leftarrow [2^n]$ — a random person
- $g_i(x) = \begin{cases} 1, & x \text{ saw movie } i; \\ 0, & \text{otherwise.} \end{cases}$
- $Y_i = g_i(X)$
- $\forall i,j$: $H(Y_i, Y_j) \le H(0.9, \frac{0.1}{3}, \frac{0.1}{3}, \frac{0.1}{3}) \le \frac{2}{3}$
- $H(Y = (Y_1, \ldots, Y_m)) \le \sum_i^{m/2} H(Y_i, Y_{i+\frac{m}{2}}) < \frac{3n}{2} \cdot \frac{2}{3} = n = H(X)$
- Hence, X is not determined by Y

Why $H(X_1,\ldots,X_n) \leq \sum_i H(X_i)$ so useful?

- $\mathcal{S} \subseteq \{0,1\}^n$; $X = (X_1,\ldots,X_n) \leftarrow \mathcal{S}$
- $\log|\mathcal{S}| = H(X) \leq \sum_i (X_i)$ implies $|\mathcal{S}| \leq 2^{\sum_i (X_i)}$
- If $\sum_i H(X_i)$ is small, then $\mathcal{S}$ is small

  X_i are unbalanced, e.g., $\sim (0.1, 0.9)$, implies $|\mathcal{S}| \leq 2^{n \cdot h(0.1)} \leq 2^{n/2}$
- $\mathcal{S}$ is large implies $\sum_i H(X_i)$ is large, hence most X_i are almost balanced
- $|\mathcal{S}| \geq 2^n/2$ implies $\mathrm{E}_{i \leftarrow [n]}[H(X_i)] \geq 1 - \frac{1}{n}$
- Most X_i are close to uniform

Hamming ball

- $p \leq \frac{1}{2}$; $\mathcal{S} = \{(a_1, \ldots, a_n) \in \{0,1\}^n \colon \sum_i a_i \leq pn\}$
- $|\mathcal{S}| = \sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k}$
- $X = (X_1, \ldots, X_n) \leftarrow \mathcal{S}$
- $\sum_i X_i \leq pn \implies \mathrm{E}\left[\sum X_i\right] \leq pn$, and by symmetry $\mathrm{E}\left[X_i\right] \leq p$ for every i
- $\forall i, j$: $\Pr[X_i = 1] = \Pr[X_j = 1] \leq p$

$\implies$ $H(X_i) \leq h(p)$ for every i

$\implies$ $|\mathcal{S}| \leq 2^{nh(p)}$

$\implies$ $\sum_{k=0}^{\lfloor pn \rfloor} \binom{n}{k} \leq 2^{nh(p)}$

- $\ldots$

Application

- $X_1, \ldots, X_n$ iid uniform bits (i.e., $\sim (\frac{1}{2}, \frac{1}{2})$)
- $\Pr\left[\sum_i X_i \leq pn\right] = \Pr\left[(X_1, \ldots, X_n) \in \mathcal{S}\right] \leq 2^{nh(p)} \cdot 2^{-n} = 2^{-n(1-h(p))}$
- Very useful inequality. No Chernoff, just IT

Isoperimetric inequality

- $\mathcal{S} \subseteq \{0,1\}^n$
- Edges of $\mathcal{S}$ — $E = \{(u,v) \in \mathcal{S} : |u - v| = 1\}$

Theorem 4

$|E| \leq \frac{1}{2} \cdot |\mathcal{S}| \cdot \log |\mathcal{S}|$

- Equality if $\mathcal{S}$ is "face" : $\mathcal{S} = \{(\mathbf{x},\mathbf{y}) : \mathbf{y} \in \{0,1\}^d\}$ for some $\mathbf{x} \in \{0,1\}^{n-d}$
- Example: $\mathcal{S}$ is a **face** of the 3-dimensional cube

  $n = 3$, $|\mathcal{S}| = 4$, implies $|E| \leq \frac{1}{2} \cdot 4 \cdot \log 4 = 4$
- E_i — edges of E in direction i $\qquad$ ($E = \dot\bigcup_{i \in [n]} E_i$)
- $X = (X_1,\ldots,X_n) \leftarrow \mathcal{S}$ and $X_{-i} = (X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$

Lemma 5

$H(X_i|X_{-i}) = \frac{2|E_i|}{|\mathcal{S}|}$

Proving Thm 4:

$$\begin{aligned}
\log|\mathcal{S}| &= H(X_1,\ldots,X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_1,X_2,\ldots,X_{n-1}) \\
&\geq H(X_1|X_{-1}) + H(X_2|X_{-2}) + \ldots H(X_n|X_{-n}) = \sum_i \frac{2|E_i|}{|\mathcal{S}|} = \frac{2|E|}{|\mathcal{S}|}. \quad \square
\end{aligned}$$

Proving Lemma 5

We prove for $i = 1$

- $\mathcal{S} \subseteq \{0,1\}^n$; $X = (X_1, \ldots, X_n) \leftarrow \mathcal{S}$
- $E = \{(u,v) \in \mathcal{S} : |u - v| = 1\}$ and E_1 contains edges of E in direction 1
- $\mathcal{S}_{-1} = \{\mathbf{y} \in \{0,1\}^{n-1} : \exists x \in \{0,1\} \text{ s.t. } (x, \mathbf{y}) \in \mathcal{S}\}$.
  ($\mathcal{S}$ projected on $(2, \ldots, n)$)
- $\mathcal{S}^e_{-1} = \{\mathbf{y} \in \{0,1\}^{n-1} : (0, \mathbf{y}), (1, \mathbf{y}) \in \mathcal{S}\}$ and $\mathcal{S}^{\neg e}_{-1} = \mathcal{S}_{-1} \setminus \mathcal{S}^e_{-1}$
- $|\mathcal{S}| = 2\left|\mathcal{S}^e_{-1}\right| + \left|\mathcal{S}^{\neg e}_{-1}\right|$
- $|E_1| = \left|\mathcal{S}^e_{-1}\right|$
- $H(X|X_{-1}) = \Pr\left[X_{-1} \in \mathcal{S}^e_{-1}\right] \cdot 1 = \frac{2|\mathcal{S}^e_{-1}|}{2|\mathcal{S}^e_{-1}| + |\mathcal{S}^{\neg e}_{-1}|} = \frac{2|E_1|}{|\mathcal{S}|}$ $\square$
- ...